

高血压患者合并重大慢性病关联规则分析*

朱碧云 王妮 黄艳群 乔博文 李博 陈卉

(首都医科大学附属北京(首都医科大学生物医学(北京市石景山区卫生(首都医科大学生物医学地坛医院 北京 100015) 工程学院 北京 100069) 信息中心 北京 100043) 工程学院 北京 100069)

〔摘要〕 基于社区公共卫生系统数据集,应用 Apriori 关联规则算法,对高血压患者合并冠心病和糖尿病的潜在因素进行关联规则挖掘。全面分析高血压患者合并其他慢病的关联因素,提高慢病管理质量与效率。

〔关键词〕 高血压;糖尿病;冠心病;关联规则挖掘;社区居民健康档案

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.11.015

Analysis of Association Rule between Hypertension Patients and Accompanied Major Chronic Diseases ZHU Biyun, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China; WANG Ni, HUANG Yanqun, School of Biomedical Engineering, Capital Medical University, Beijing 100069, China; QIAO Bowen, LI Bo, Beijing Shijingshan District Health Information Center, Beijing 100043, China; CHEN Hui, School of Biomedical Engineering, Capital Medical University, Beijing 100069, China

〔Abstract〕 The paper explores the association rules of the potential factors between hypertension patients and accompanied coronary heart disease and diabetes based on the data set of the community public health system and by use of the Apriori association rule algorithm, and comprehensively analyzes the associated factors between hypertension patients and other accompanied chronic diseases, in order to improve the quality and efficiency of chronic disease management.

〔Keywords〕 hypertension; diabetes; coronary heart disease; association rule mining; health records of community residents

1 引言

高血压是临床中比较常见的一种心血管病症^[1],其病程长且无法治愈,对人们的健康有着极大的影响。流行病学调查发现近些年来高血压发病率逐年上升^[2],给患者及社会带来了巨大的经济负

担。有研究表明一些慢性病常常相互伴随发生,如高血压、肥胖、糖尿病、血脂异常等^[3-5],而且患者的性别、年龄、婚姻状况等人口学因素^[6-7]以及吸烟、饮酒、锻炼、饮食等生活习惯^[2,8]都对慢性病的发生有一定影响。因此对高血压患者的共患病现象和其他潜在关联因素进行分析将对高血压的预防和治疗有借鉴意义。

关联规则挖掘是一种数据挖掘技术^[9],用于探索事物之间的关联性,目前已被用于中医中药、临床医学、流行病学等方面的研究,如发现脑卒中危险因素^[10]、住院跌倒患者的防范^[11]、患者就诊规律挖掘^[12]、慢性阻塞性肺疾病的协同疾病^[13]、儿童安全用药^[14]、中医药处方配伍规律^[15]等。本研究利用北京市石景山区社区公卫系统数据集,通过

〔收稿日期〕 2019-05-22

〔作者简介〕 朱碧云,助理工程师,发表论文6篇;通讯作者:陈卉,教授,博士生导师。

〔基金项目〕 国家自然科学基金项目“面向跨领域异构数据的患者相似性学习方法及应用”(项目编号:81671786)。

关联规则挖掘, 探讨高血压与糖尿病、冠心病之间的关联性及其关联强度, 分析其他潜在关联因素, 研究高血压患者的共患病现象, 提供有效防治建议。

2 资料和方法

2.1 数据来源

为 2018 年北京市石景山区社区卫生系统数据集, 其中包括人口学数据 (性别、年龄、婚姻状况等)、生活习惯 (锻炼、饮酒、吸烟、饮食等)、家族史、慢性病患病情况 (糖尿病、冠心病、脑卒中)、生活条件 (饮用水、燃料类型、厨房排风设施等) 等 106 个指标。为了简化计算并尽可能挖掘出有效的关联规则, 本研究从中提取性别、年龄 (>60 岁定义为老年人)、婚姻状况、身高、体重、是否患高血压、是否患糖尿病、是否患冠心病、是否患脑卒中、父亲有无高血压、母亲有无高血压、锻炼 (经常锻炼、锻炼无规律、不锻炼)、饮酒 (经常饮酒、饮酒无规律、不饮酒)、吸烟 (经常吸烟、已戒烟、不吸烟)、睡眠 (睡眠困难、入睡困难、早醒、梦游、嗜睡等)、饮食习惯 (偏咸、偏甜、偏油、嗜热食、素食、辛辣等) 共 16 个指标。此外根据身高和体重计算体重指数 BMI, 参考世界卫生组织的标准, 将 BMI >24 kg/m² 定义为肥胖。

2.2 关联规则挖掘算法

关联规则是数据挖掘中一项重要研究内容, 用于发现数据背后以及事物之间可能存在的关联或联系, 属于一种无监督学习。若两个或多个变量的取值之间存在某种规律性就称为关联。对于数据库中的 1 条记录, 若其同时具有特征属性 A 与特征属性 B, 则称特征属性 A 和 B 是关联的, 即构成 1 条关联规则 A → B。其中, A 和 B 分别称为关联规则的先导和后继。A 和 B 不一定是因果关系, 它们很大程度上为“共生关系”。判断关联规则强弱的指标主要有以下 3 个: 一是支持度 Support (A → B) = P (AB), 即在样本数据中 A、B 同时发生的频率; 二是置信度 Confidence (A → B) = P (B | A), 即在样本数据中 A 发生的前提下 B 发生的频

率; 三是提升度 Lift (A → B) = P (B | A) / P (B), 即在样本数据中两种情况下频率的比值。进行关联规则挖掘最基本也是最著名的是 Apriori 算法, 其核心思想是基于频集理论的一种递推方法, 从数据库中挖掘那些支持度和置信度都不低于给定最小支持度阈值和最小置信度阈值的关联规则。

2.3 分析方法

将原始数据导入 Access 软件, 提取所有高血压患者的相应分析指标作为研究内容建立数据集并进行数据清理。采用 Python (<https://www.python.org/>) 数据挖掘软件进行 Apriori 算法分析, 挖掘支持度 >10%、置信度 >5% 的规则并定义提升度 >1 的规则为强关联规则。采用 Gephi (<https://gephi.org/>) 软件绘制关联规则的网状图, 以节点表示特征, 用节点间连线的粗细直观地展示特征间关联性的强弱。

3 结果

3.1 高血压患者合并慢病情况

从总数据集中共确定 73 894 名高血压患者, 年龄为 67.5 ± 12.0 岁。所有高血压患者中, 55 649 人 (75.3%) 至少合并冠心病或糖尿病之一。其中 32 992 人 (44.6%) 合并冠心病, 28 427 人 (38.5%) 合并糖尿病, 16 185 人 (21.9%) 则同时合并这两种慢病。男性高血压患者合并冠心病和糖尿病的发生率 (42.7% 和 38.4%) 高于女性 (38.9% 和 35.8%, P 值均 < 0.001), 老年患者 (49.8% 和 39.8%) 高于非老年患者 (38.0% 和 36.7%, P 值均 < 0.001), 见表 1。

表 1 高血压合并冠心病、糖尿病患者的性别、年龄分布 (n [%])

项目	总人数 (n = 73 894)	合并冠心病 (n = 32 992)	合并糖尿病 (n = 28 427)
性别	男	32 932 (44.6)	14 048 (42.6)
	女	10 753 (14.6)	4 186 (12.7)
	不详	30 209 (40.9)	14 758 (44.7)
年龄	≤60 岁	32 150 (43.5)	12 213 (37.0)
	>60 岁	41 744 (56.5)	20 779 (63.0)

3.2 高血压合并糖尿病关联规则

表 2 列出了有 1 个或两个条件（先导）的部分强关联规则（提升度分别 >1 和 1.25）。从第 1 条规则可知高血压患者同时合并冠心病和糖尿病的概率是 21.9%；在所有合并冠心病的患者中有将近一半（49.1%）的人还患有糖尿病；高血压患者合并冠心病使其合并糖尿病的风险增加 30%。此外年龄、饮食习惯和锻炼情况都与糖尿病有关联性。

表 2 高血压合并糖尿病的关联规则挖掘结果

规则	支持度 (%)	置信度 (%)	提升度
单一 冠心病→糖尿病	21.9	49.1	1.28
先导 饮食偏咸→糖尿病	8.8	40.6	1.05
的规 肥胖→糖尿病	14.9	40.2	1.04
则 老年人→糖尿病	22.5	39.8	1.03
锻炼→糖尿病	24.6	39.8	1.03
锻炼不规律→糖尿病	6.1	39.6	1.03
两个 冠心病, 肥胖→糖尿病	8.5	51.0	1.32
先导 冠心病, 非老年人→糖尿病	8.4	50.9	1.32
的规 冠心病, 饮食偏咸→糖尿病	5.2	49.8	1.30
则 冠心病, 锻炼→糖尿病	13.7	49.6	1.29
冠心病, 已婚→糖尿病	17.8	49.6	1.29
冠心病, 有睡眠障碍→糖尿病	19.8	49.5	1.29
冠心病, 不饮酒→糖尿病	16.7	49.4	1.28
冠心病, 饮食不偏甜→糖尿病	20.9	49.3	1.28
冠心病, 不吸烟→糖尿病	16.6	49.3	1.28
冠心病, 男→糖尿病	9.4	49.2	1.28
冠心病, 饮食不偏咸→糖尿病	16.3	49.2	1.28
冠心病, 母亲无高血压→糖尿病	19.2	48.6	1.26
冠心病, 体重正常→糖尿病	12.8	48.5	1.26
冠心病, 父亲无高血压→糖尿病	18.9	48.3	1.25

3.3 高血压合并冠心病关联规则

表 3 列出了包含有 1 个或两个条件（先导）的部分强关联规则（提升度分别 >1 或 1.25）。结果显示高血压合并糖尿病的患者其冠心病的发生率达 56.9%，而合并糖尿病且有睡眠障碍使冠心病的发生风险提高到 60.4%。此外年龄、家族史、饮食习惯和锻炼情况都与冠心病有关联性。表 2 和表 3 列出的强关联规则中所有单一先导规则的关联网络，见图 1。可见高血压患者合并糖尿病与合并冠心病

间有很强的关联性，饮食偏咸、锻炼不规律与两种慢病都有关联，老年高血压患者常常伴随发生糖尿病与冠心病。

表 3 高血压合并冠心病的关联规则挖掘结果

规则	支持度 (%)	置信度 (%)	提升度
单一 糖尿病→冠心病	21.9	56.9	1.28
先导 老年人→冠心病	28.1	49.8	1.12
的规 父亲有高血压→冠心病	5.6	49.1	1.10
则 锻炼不规律→冠心病	7.5	48.1	1.08
有睡眠障碍→冠心病	40.0	48.0	1.08
饮食偏咸→冠心病	10.4	47.7	1.07
母亲有高血压→冠心病	5.1	46.0	1.03
两个 糖尿病, 有睡眠障碍→冠心病	19.8	60.4	1.35
先导 糖尿病, 老年人→冠心病	13.5	60.0	1.35
的规 糖尿病, 饮食偏咸→冠心病	5.2	58.6	1.31
则 糖尿病, 体重正常→冠心病	12.8	57.4	1.29
糖尿病, 饮食不偏甜→冠心病	20.9	57.3	1.28
糖尿病, 不饮酒→冠心病	16.7	57.3	1.28
糖尿病, 不吸烟→冠心病	16.6	57.2	1.28
糖尿病, 饮食不偏咸→冠心病	16.3	57.0	1.28
糖尿病, 肥胖→冠心病	8.5	56.8	1.27
糖尿病, 母亲无高血压→冠心病	19.2	56.6	1.27
糖尿病, 已婚→冠心病	17.8	56.5	1.27
糖尿病, 父亲无高血压→冠心病	18.9	56.2	1.26
糖尿病, 锻炼→冠心病	13.7	55.9	1.25

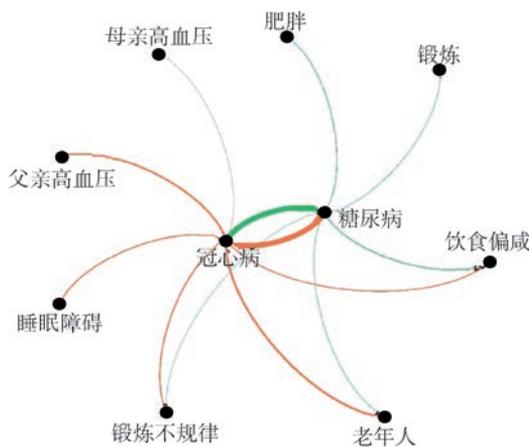


图 1 关联规则分析网络

注：图中线的粗细与规则提升度成比例

4 讨论

4.1 高血压患者合并糖尿病关联因素

目前高血压合并糖尿病的患者越来越多。本研

究的人群中高血压患者的糖尿病患病率达 38.5%，远高于 2004 年文献报道的 10%~20%^[16]。高血压合并糖尿病后，患者控制血糖和血压水平更加不易，使得脑卒中、冠心病等严重心脑血管疾病的发生甚至死亡风险增加^[17-18]。在对高血压患者合并糖尿病的关联规则分析中，冠心病→糖尿病、饮食偏咸→糖尿病、肥胖→糖尿病是关联性最强的 3 条单前导规则，伴冠心病、饮食偏咸以及肥胖使高血压患者合并糖尿病的风险分别增加 28%、5% 和 4%。而两前导关联规则挖掘结果进一步显示，对于高血压伴冠心病患者，肥胖和饮食偏咸又使患者再合并糖尿病的可能性分别增加 1.9% 和 0.7%。从这些强关联规则可以看出，高血压患者伴冠心病增加糖尿病患病可能性，肥胖、饮食偏咸则使这种可能性进一步增大。其他研究也显示在 65 岁以上高血压患者中肥胖是高血糖的重要危险因素^[19]。

4.2 高血压患者合并冠心病关联因素

冠心病是目前世界范围内最首要的疾病死亡原因^[20]。高血压是冠心病发生的重要危险因素，控制高血压不仅是预防冠心病的关键，也是减轻心血管疾病所造成的沉重经济负担和保证患者生命及生活质量的关键^[21]。临床和流行病学研究显示中老年人高血压伴糖尿病与冠心病存在密切联系，冠心病的发生率显著提高^[22-23]。本研究发现伴糖尿病使高血压患者合并冠心病的可能性增加 12.3%，与其他研究结果一致。高血压合并糖尿病的患者冠状动脉病变程度更重，临床预后差^[24]，因此积极治疗高血压和糖尿病将有利于冠心病的防治。本研究人群中睡眠障碍导致高血压伴糖尿病患者发生冠心病的可能性增加 3.5%。其他研究揭示睡眠与高血压、糖尿病和冠心病的发病都具有一定的关联性，睡眠时间过短或过度睡眠均会提升高血压、糖尿病和冠心病的发病率^[25]。高盐饮食可进一步导致高血压合并糖尿病患者发生不良心血管事件^[24]，与本研究发现饮食偏咸会导致高血压伴糖尿病患者发生冠心病的风险增加（从 56.9% 提高到 58.6%）一致。此外年龄一直以来都是重大慢性疾病的危险因素，高血

压伴冠心病的老年患者的糖尿病患病率提高了 1.8%，高血压伴糖尿病老年患者的冠心病患病率提高了 3.1%，社区高血压管理的重点仍将放在中老年人群上。

5 结语

本研究利用数据挖掘技术中的关联规则挖掘方法，基于社区居民健康档案数据，对近 74 000 名社区高血压患者合并慢病及其影响因素进行分析。研究结果提示社区慢病管理中应倡导健康生活方式（如高质量睡眠、适当锻炼、低盐饮食、控制体重），定期监测高血压、糖尿病、冠心病的发生和变化，避免共患病的出现。通过关联规则挖掘，全面分析高血压患者合并其他慢性病的关联因素，使得相关疾病的预防和干预更具有针对性，同时也能提高慢病管理效率。

参考文献

- 1 陈伟. 吸烟与高血压患者心血管疾病死亡和全因死亡关系的前瞻性队列研究 [J]. 实用心脑血管病杂志, 2015, 23 (11): 5-9.
- 2 姚杏娟, 朱一胜, 丁益民, 等. 社区人群吸烟和肥胖交互作用与高血压的关联分析 [J]. 南京医科大学学报, 2014, 34 (7): 955-958.
- 3 李生梅, 李满桂, 贾军, 等. 腹型肥胖人群、血脂、血管紧张素 II 与高血压病的探讨 [J]. 中国临床医生杂志, 2015, 43 (10): 29-31.
- 4 任仙龙, 胡冬梅, 王文娟. 关联规则在社区居民慢性病患病率分析中的应用 [J]. 中国卫生统计, 2013, 30 (6): 818-820.
- 5 姚惠芳, 尹冬彩, 胡建强. 肥胖高血压糖尿病高血脂与冠心病的相关分析 [J]. 浙江中西医结合杂志, 2004, 14 (4): 259.
- 6 刘昊, 李骏, 张庆军. 湖北省职业人群不同性别慢性病高风险人群健康危险因素分析 [J]. 中国慢性病预防与控制, 2016, 24 (9): 681-683.
- 7 吕鹏举, 原杰, 宁梓滢, 等. 大庆油田职工慢性病现状及年龄预警模型分析 [J]. 现代生物医学进展, 2016, 16 (14): 2751-2754.
- 8 时松和, 郭朝阳, 范军星. 河南省 60 岁以上城市居民

- 慢性病患病特征分析 [J]. 现代预防医学, 2016, 43 (1): 115 - 118.
- 9 Agrawal R, Imielinski T, Swami A. Mining Association Rules between Sets of Items in Large Databases [C]. Washington D. C; Proc. ACM SIGMOD Int Conf. on Management of Data, 1993: 207 - 216.
- 10 罗一夫, 何健, 赵会晶. 基于 Apriori 关联规则的脑卒中危险因素分析 [J]. 中国数字医学, 2017, 12 (11): 90 - 93.
- 11 沈鸣雁, 王华芬, 封秀琴, 等. 住院跌倒患者的数据挖掘与跌倒防范对策分析 [J]. 中华护理杂志, 2017, 52 (9): 1087 - 1091.
- 12 翟运开, 武戈. 基于电子病历信息大数据挖掘的患者就医行为分析 [J]. 医学信息学杂志, 2017, 38 (7): 12 - 17.
- 13 贺建峰, 姚旭升, 杨静, 等. 基于关联规则算法的慢性阻塞性肺疾病及其协同发病分析 [J]. 北京生物工程, 2018, 37 (6): 72 - 77, 84.
- 14 胡珊珊, 厉玉鹏. 关联规则技术在儿童安全用药中的应用 [J]. 医学信息学杂志, 2018, 39 (10): 73 - 77.
- 15 张庆, 轩扬, 马恒毅. 关联规则在中医药方剂挖掘中的应用 [J]. 医学信息学杂志, 2018, 39 (11): 72 - 75.
- 16 于汇民, 刘国仗. 高血压与糖尿病共患与心血管疾病的关系 [J]. 中国分子心脏病学杂志, 2004, 4 (1): 52 - 55.
- 17 苏馨. 高血压合并糖尿病血压控制的探讨 [J]. 重庆医学, 2018, 47 (16): 2219 - 2222.
- 18 刘国仗, 于汇民. 高血压合并糖尿病的治疗 [J]. 中国心血管杂志, 2002, 7 (2): 84 - 89.
- 19 高红旭, 张良, 范景丽, 等. 老年高血压患者血糖、血脂与体质指数的相关性分析 [J]. 中国基层医药, 2019, 26 (5): 513 - 517.
- 20 GBD 2016 Causes of Death Collaborators. Global, Regional, and National Age - sex Specific Mortality for 264 Causes of Death, 1980 - 2016: a systematic analysis for the Global Burden of Disease Study 2016 [J]. Lancet, 2017, 390 (10100): 1151 - 1210.
- 21 翟屹, 胡建平, 孔灵芝, 等. 中国居民高血压造成冠心病和脑卒中的经济负担研究 [J]. 中华流行病学杂志, 2006, 27 (9): 744 - 747.
- 22 梁德贤, 陈康荣, 李庆军. 中老年高血压伴糖尿病与冠心病相关性的研究 [J]. 数理医药学杂志, 2018, 31 (8): 1149 - 1150.
- 23 任进平, 李松, 宋媛媛. 中老年高血压伴糖尿病与冠心病相关性的探讨 [J]. 中国医药导刊, 2016, 18 (12): 1237 - 1238.
- 24 张艳, 赵昕, 张效林, 等. 高血压前期合并糖尿病患者临床特点及高盐饮食对其预后影响 [J]. 临床军医杂志, 2018, 46 (3): 241 - 247.
- 25 董红, 李连海. 睡眠时间与高血压、糖尿病和冠心病关系的现况研究 [J]. 临床医药文献电子杂志, 2017, 4 (78): 15251 - 15254.

(上接第 65 页)

- 8 Morton M E. Use and Acceptance of an Electronic Health Record: factors affecting physician attitudes [D]. Oxford: University of Mississippi, 2008.
- 9 Hodgson L, Aiken P. Organizational Change Enabled by the Mandated Implementation of New Information Systems Technology: a modified technology acceptance model [C]. Boston; ACM Sigcpr Conference on Computer Personnel Research, 1998: 205 - 213.
- 10 Aggelidis V P, Chatzoglou P D. Using a Modified Technology Acceptance Model in Hospitals [J]. International Journal of Medical Informatics, 2009, 78 (2): 115 - 126.
- 11 Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions [J]. Decision Sciences, 2010, 39 (2): 273 - 315.
- 12 李楠, 郜凯华, 谢铮. 医院组织因素对实施电子病历的影响综述 [J]. 学理论, 2013 (35): 63 - 65.
- 13 郜凯, 李楠, 谢铮. 组织因素对实施电子病历系统的影响 [J]. 医院管理论坛, 2013 (11): 26 - 28.
- 14 Bhattacharjee A, Hikmet N. Physicians' Resistance toward Healthcare Information Technology: a theoretical model and empirical test [J]. European Journal of Information Systems, 2007, 16 (6): 725 - 737.
- 15 Boonstra A, Broekhuis M. Barriers to the Acceptance of Electronic Medical Records by Physicians from Systematic Review to Taxonomy and Interventions [J]. BMC Health Services Research, 2010, 10 (1): 231.
- 16 Lin C, Lin I C, Roan J. Barriers to Physicians' Adoption of Healthcare Information Technology: an empirical study on multiple hospitals [J]. Journal of Medical Systems, 2012, 36 (3): 1965 - 1977.